Management Science

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Two New Features in Discrete Choice Experiments to Improve Willingness-to-Pay Estimation That Result in SDR and SADR: Separated (Adaptive) Dual Response

Christian Schlereth, Bernd Skiera

informs
http://pubsonline.informs.org/journal/mnsc/

MANAGEMENT SCIENCE
Vol. 63, No. 3, March 2017, pp. 829–842
ISSN 0025-1909 (print), ISSN 1526-5501 (online)

Two New Features in Discrete Choice Experiments to Improve Willingness-to-Pay Estimation That Result in SDR and SADR: Separated (Adaptive) Dual Response

Christian Schlereth,[a] **Bernd Skiera**[b]

[a] WHU – Otto Beisheim School of Management, Marketing and Sales Group, Chair of Digital Marketing, 56179 Vallendar, Germany;
[b] Faculty of Business and Economics, Department of Marketing, Goethe-University Frankfurt, 60323 Frankfurt am Main, Germany

Contact: christian.schlereth@whu.edu (CS); skiera@skiera.de (BS)

Received: December 9, 2013
Revised: November 19, 2014; July 31, 2015
Accepted: September 19, 2015
Published Online in Articles in Advance: March 22, 2016

https://doi.org/10.1287/mnsc.2015.2367

Abstract. Discrete choice experiments such as choice-based conjoint and dual response are popular preference elicitation methods, yet, they can suffer from context effects, extreme response behavior, and problems with estimating consumers' willingness to pay accurately when the purchase probability is high. This study proposes two new features to avoid these limits: (a) strictly separating all forced and all free choice questions and (b) an adaptive mechanism to select fewer, but more informative, free choice questions. The use of these two features invokes two new discrete choice experiment methods: separated dual response (SDR), with just the first feature, and separated adaptive dual response (SADR), with both features. A conceptual comparison, simulation study, tests to address endogeneity concerns, and three empirical studies demonstrate the appeal of these two new features, relative to the benefits of existing discrete choice experiments, especially when estimating willingness to pay.

History: Accepted by Pradeep Chintagunta, marketing.
Supplemental Material: Data and the online appendix are available at https://doi.org/10.1287/mnsc.2015.2367.

Keywords: willingness to pay • discrete choice experiments • choice-based conjoint • pricing • no-purchase option

1. Introduction

Companies and researchers alike use choice-based conjoint and dual response to measure consumers' preferences and willingness to pay (WTP).[1] These methods differ primarily in their use of free and forced choice questions. Forced choice questions only contain product alternatives; free choice questions include product alternatives and a no-purchase option. Whereas choice-based conjoint usually contains only free choice questions, dual response contains sequences of one forced and one free choice question. In the free choice question, the respondent must choose between the product alternative chosen in the previous forced choice question and a no-purchase option.

Valid results in choice-based conjoint and dual response require the exclusion of context effects and extreme response behavior, two requirements that relate to how respondents use the no-purchase option. To avoid context effects, a respondent's sole reason for selecting the no-purchase option must be that none of the product alternatives in the choice set provides sufficient utility to justify a purchase. Avoiding extreme response behavior requires that researchers exclude conditions in which consumers always or never choose the no-purchase option.

Yet, these requirements are typically not fulfilled; choice-based conjoint and dual response still frequently suffer from context effects (Dhar 1997, Dhar and Simonson 2003, Rooderkerk et al. 2011) and extreme response behavior (Gensler et al. 2012). Another problem with discrete choice experiments arises because respondents' purchase probability affects the accuracy of estimates of their WTP. As we show, these methods measure WTP less accurately if respondents have a high probability of buying, compared with respondents who exhibit a low purchase probability.

Therefore, we seek to avoid these shortcomings in three ways. First, we suggest two new features that might improve the validity of WTP estimates. Second, we apply these two features to develop two new methods for use in discrete choice experiments: separated dual response (SDR), which overcomes context effects by imposing a strict separation between all forced and free choice questions, and separated adaptive dual response (SADR), which imposes the same separation and also features an adaptive mechanism to avoid extreme response behavior and ensure that accurate estimates of WTP do not depend on respondents' probability of buying. Third, in three empirical studies, we compare the proposed new methods with choice-based conjoint and dual response, with a particular focus on SADR as the most elaborated method.

The new methods differ in several notable ways from other adaptive discrete choice experiments, such as those proposed by Toubia et al. (2007) and Yu et al. (2011). The adaptive part of SADR focuses on gathering information about the threshold that makes each respondent indifferent between buying and not buying, whereas other adaptive discrete choice experiments seek information about each respondent's preferences for attributes and their levels. An exception is the individually adapted choice-based conjoint (IACBC) of Gensler et al. (2012), which systematically varies prices in the choice sets, such that prices continuously shift upward every time the respondent selects a product alternative and downward when the respondent selects the no-purchase option. However, similar to choice-based conjoint, IACBC remains susceptible to context effects with respect to the no-purchase option, and, as we show, its adaptation of prices introduces behavioral endogeneity.

2. Review of Discrete Choice Experiments

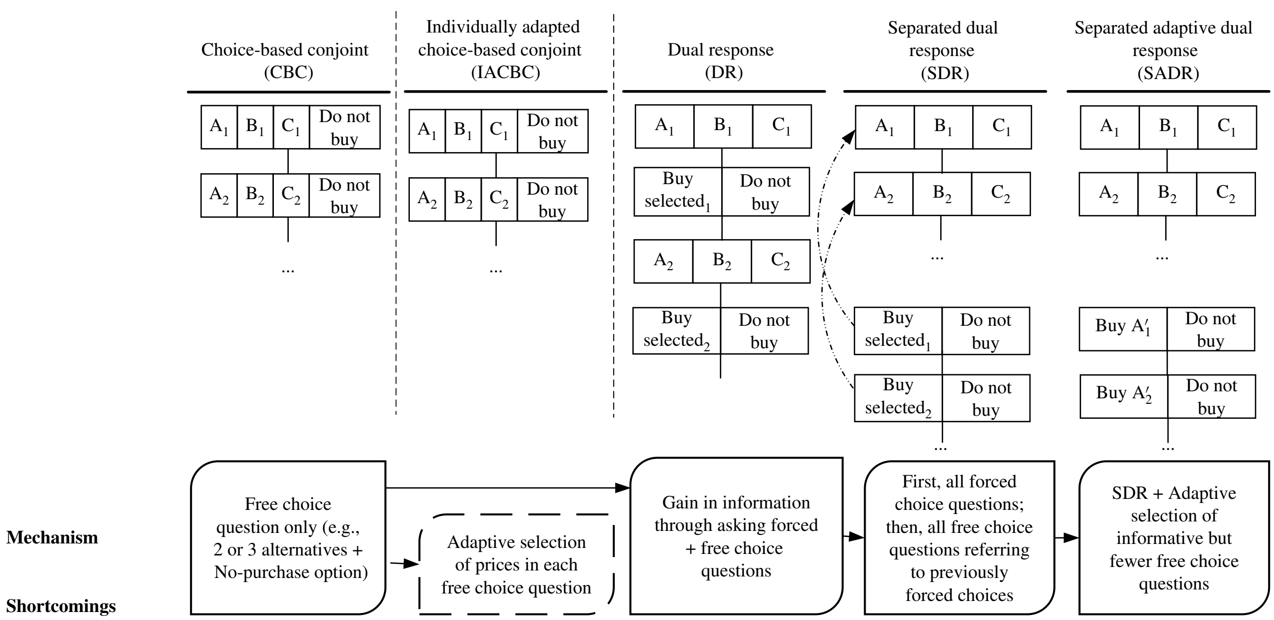
Figure 1 details existing discrete choice experiments, which use different types and sequences of questions.[2] They each suffer from at least two of four shortcomings: context effects, extreme response behavior, effect of respondents' purchase probability on accuracy of WTP estimates, and behavioral endogeneity.

2.1. Choice-Based Conjoint and Its Shortcomings

Choice-based conjoint usually contains only free choice questions, consisting of several product alternatives and a no-purchase option (e.g., Brazell et al. 2006, Feit et al. 2010, Louviere et al. 2000, Miller et al. 2011). Despite its popularity, robust evidence shows that repeatedly asking respondents to make decisions among choice sets suffers from context effects with respect to the no-purchase option (e.g., Dhar 1997, Rooderkerk et al. 2011), such as an attraction effect. This context effect occurs if a dominant product alternative exists in a choice set, because its dominance decreases the likelihood that the respondent selects the no-purchase option. Another context effect arises when a choice set contains similarly attractive product alternatives, in which case the decision to select one of them is more difficult, so respondents select the no-purchase option as an "easy way out" (Dhar 1997).

Choice-based conjoint also suffers from extreme response behavior, that is, when respondents always or never choose the no-purchase option (Gensler et al. 2012), which gives the researcher no information about the conditions in which respondents will stop or start buying a product, such that inferences about their WTP are not possible. In Gensler et al. (2012), 58% of

Figure 1. Comparison of Discrete Choice Experiments



	CBC	IACBC	DR	SDR	SADR
Context effects	X	X	X	—	—
Extreme response behavior	X	—	X	X	—
Impact of purchase probability on accuracy of WTP	X	—	X	X	—
Behavioral endogeneity concerns	—	X	—	—	—

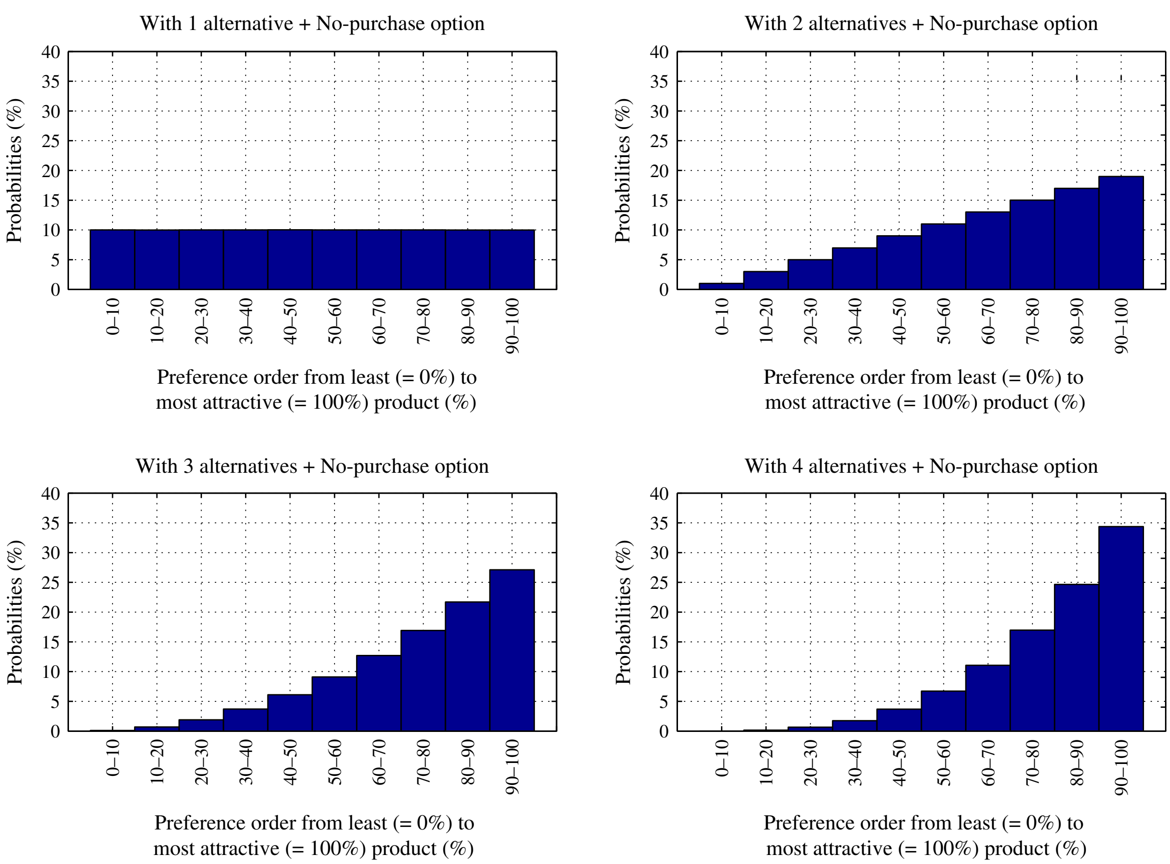
respondents exhibited such extreme response behavior. Parker and Schrift (2011) similarly report that 63.5% of their respondents never picked the no-purchase option, so the results of the experiment could not offer estimates of the WTP for more than half of the respondents.

Another shortcoming of choice-based conjoint is related to the widely accepted recommendation that choice sets should cover a fairly large range of prices (and other attribute levels). Thus, prices are higher than in reality but not so high as to be unbelievable (Rao 2014, p. 44). Adherence to this convention can lead to systematic differences in the WTP estimate accuracy if multiple product alternatives appear in a choice set, however. Specifically, when deciding whether to select a product or the no-purchase option, the respondent compares the no-purchase option with the most attractive product alternative in the choice set. If multiple product alternatives with different levels of attractiveness appear in the choice set, it likely contains not just products with "average" attractiveness but also more attractive product alternatives. In turn, WTP estimates will be better for respondents with lower probabilities of buying a product from the category than for those with higher probabilities.

To illustrate this effect, consider a simulation of the attractiveness of the most preferred product alternatives in choice sets of different sizes (1–4 product alternatives + no-purchase option). The commonly applied D-efficient design generation technique systematically varies attribute levels to form product alternatives in choice sets but ignores the existence of the no-purchase option. Therefore, any given choice set contains product alternatives with random degrees of attractiveness to the respondent. To imitate such randomness, we select products for the choice set by randomly drawing values of the attractiveness of various products from a uniform distribution [0%; 100%], where values close to 0% indicate unattractive but values close to 100% signal more attractive product alternatives. The highest randomly drawn value is considered to be the most attractive product alternative, against which we can compare the no-purchase option for this simulation.

Figure 2 depicts the likelihood that a product with a given level of attractiveness is the most attractive one in choice sets of varying sizes. The left-hand side of Figure 2 offers deciles with the least attractive products; the right-hand side shows the most attractive products. When the choice set includes just one

Figure 2. (Color online) Distribution of Probabilities That Products in Different Deciles Compare with No-Purchase Option



product alternative, products in all deciles have the same probability of being compared against the no-purchase option. Adding more product alternatives then increases the likelihood that more attractive products are compared against the no-purchase option. For example, the probability of comparing an alternative with less-than-average attractiveness with the no-purchase option decreases from 50% in the case of one product alternative to 25%, 12.5%, and 6.3% for choice sets with two, three, and four product alternatives, respectively. A respondent with a high purchase probability, by definition, should prefer more attractive product alternatives to the no-purchase option. Thus, this respondent's choices are less informative than those of a respondent with a lower purchase probability, whose preference for the more attractive product alternatives is closer to that for the no-purchase option. A respondent's high purchase probability lowers accuracy of WTP estimates.

2.2. Dual Response and Its Shortcomings

Dual response contains repeated sequences of one forced and one free choice question. The forced choice question contains several product alternatives; the free choice question asks the respondent to choose between a selected product alternative (from the forced choice question) and a no-purchase option. These sequences avoid the context effects of choice-based conjoint. However, they also introduce a new context effect, choice deferral, such that respondents tend to select the no-purchase option up to 36% more often (Dhar and Simonson 2003; see also Brazell et al. 2006 and Wlömert and Eggers 2016). Dhar and Simonson (2003) attribute this effect to the negative emotions that respondents associate with a forced choice decision, which they mitigate by not committing to a previous choice. Similar to choice-based conjoint, dual response suffers from a negative impact of respondents' purchase probability on WTP estimation accuracy, because respondents still compare the no-purchase option against the most attractive product alternative in a choice set.

The forced choice questions in dual response aim to reduce the consequences of extreme response behavior, especially among respondents who often or always choose the no-purchase option; these choices would have provided no information about their preferences for particular product attributes in choice-based conjoint. Yet, these respondents also face twice as many questions as they would in a choice-based conjoint study, which increases both cognitive effort and error variances (Bech et al. 2011).

Dual response contains large potentials for reducing the cognitive effort, because some free choice questions are redundant because they do not build on previous answers. For example, if a respondent already plans to purchase product A for $10, there is no need to ask if that respondent would purchase a better product at the same or a lower price. Similarly, if a respondent would not purchase product B for $15, there is no need to ask if that respondent would purchase a worse product at the same or a higher price.

Finally, repeatedly asking alternating questions—forced and then free choices—requires respondents to perform two different, alternating, decision-making tasks, which demands cognitive effort and might cause fatigue or annoyance. This effect and the greater number of questions tend to hinder the validity of the results (Dhar and Simonson 2003). In two empirical studies, Brazell et al. (2006) thus show that dual response is not superior to choice-based conjoint.

2.3. IACBC and Its Shortcomings

Gensler et al. (2012) developed IACBC to address extreme response behavior in choice-based conjoint. This method presents free choice questions, consisting of multiple product alternatives and a no-purchase option. To reduce the likelihood of extreme response behavior, IACBC adaptively manipulates prices in the free choice questions upward when the respondent selects a product alternative and downward if the respondent selects the no-purchase option. Consequently, it is susceptible to context effects with respect to the no-purchase option, similar to choice-based conjoint, because it always invokes a comparison against multiple product alternatives. As we show subsequently, its adaptation of prices also introduces behavioral endogeneity, such that respondents' WTP decreases in the process of responding to the choice sets.

3. Two New Features and Implementations in Two New Discrete Choice Experiments

We propose two new features to overcome these shortcomings and implement them in two discrete choice experiments. First, we strictly separate all forced from all free choice questions, which we implement in separated dual response (SDR). Second, we also use an adaptive mechanism to select fewer, but more informative, free choice questions, which is implemented in separated adaptive dual response (SADR).

3.1. Separated Dual Response (SDR) with the Separation Feature

The first new feature, strictly separating forced and free choice questions, builds on Dhar and Simonson's (2003) efforts to avoid choice deferral in a single choice set. They argue that choice deferral in dual response occurs because respondents try to mitigate the negative emotions associated with forced choice questions; they demonstrate empirically that the effect diminishes with more time between the forced and free choice questions. We transfer this notion from a single to multiple choice sets: Instead of asking sequences of forced and free choice questions, as in dual response, we propose asking all forced choice questions first, and only thereafter asking any free choice question.

This new method, which we call separated dual response (SDR), is similar to dual response but uses a distinct separation feature. By putting all forced choice questions into one block and all free choice questions in another block, it requires respondents to consider the block with the forced choice questions first, before they decide whether to purchase each of the chosen product alternatives or not. Thus, SDR is flexible enough to support any design of the forced-choice block, such as D- or A-optimal, Bayesian, or even adaptive designs (Toubia et al. 2007, Yu et al. 2011). That is, dual response and SDR use the same questions but in different orders.

3.2. Separated Adaptive Dual Response (SADR) with the Separation and Adaptive Selection Features

Similar to SDR, separated adaptive dual response (SADR) puts all forced and free choice questions into separate blocks, but it also uses an adaptive procedure to select different, fewer, but more informative free choice questions. As we illustrate in Figure 3, SADR consists of six steps.

The central idea is to use information from the decisions in the forced choice block (step 1) to estimate a preliminary utility function for each respondent in step 2, which then reveals the preliminary order for all feasible attribute-level combinations in step 3. It subsequently selects from this order the most informative products for the free choice block (steps 4 and 5). After acquiring the anticipated number of completed questionnaires, the researcher can estimate the parameters of the final utility function (step 6), using a scale-extended hierarchical Bayes sampler.

In step 1, SADR is flexible enough to employ any design in the forced-choice block. The estimation (step 2) must be rapid to avoid any noticeable time delays between presentation of the forced choice questions and the free choice questions. Sampling approaches that rely on hierarchical Bayes are too slow. Traditional maximum likelihood methods also fail because they lack a sufficient number of observations. Instead, the estimation approach is motivated by Heckman and Snyder (1997) who propose the use of a linear probability model that is fast and allows for a reasonable accurate approximation of true preferences. The linear probability model is simple for both estimation and interpretation. However, it suffers from nonnormal errors and predictions that fall outside the unit interval [0; 1], which mainly affect predictions of the precise probability of choosing a certain product, which is not the goal of this step. Instead, SADR seeks only to determine the order of choice probabilities for all feasible attribute-level combinations and rank all possible product alternatives from least to most attractive (step 3). We refer to this order as the preliminary preference order.[3]

In steps 4 and 5, the free choice block shows $n \times m$ products and asks—in m iterations (step 5), each consisting of n free choice questions (step 4)—whether the respondent would buy or not buy the presented product. Each free choice question consists of one product and the no-purchase option. In step 4, SADR selects the product alternatives for the first n (e.g., 3) free choice questions, which might differ from those shown in the forced choice block. It chooses those n products that divide the whole preliminary preference order

Figure 3. Six-Step Separated Adaptive Dual Response (SADR) Procedure

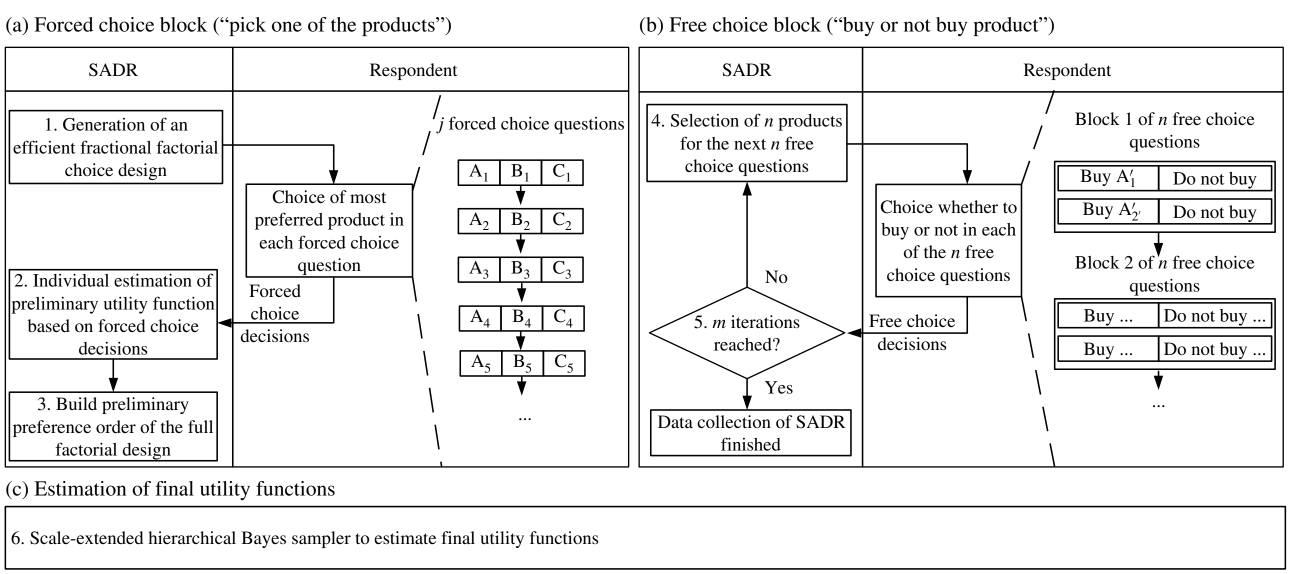
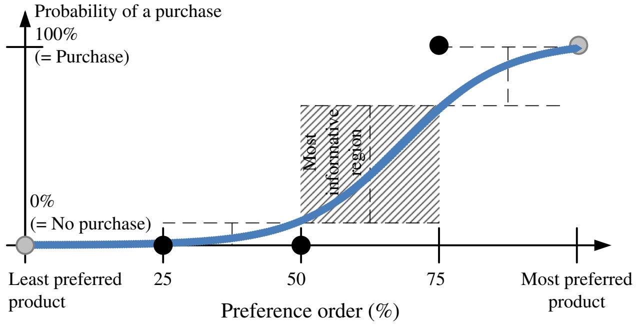


Figure 4. (Color online) Most Informative Region for Separated Adaptive Dual Response (SADR)



Note. The most informative region is the area with the largest difference in purchase probabilities.

between the least and most preferred product into $(n+1)$ equal-sized ranges (see Figure 4). For example, if a discrete choice experiment consists of 4 attributes with 4 levels each, its preliminary preference order consists of $4^4 = 256$ products (where position 256 indicates the position of the most preferred product). Then, SADR selects products at positions 64 (~25% in the preliminary preference order), 128 (~50%), and 192 (~75%) and presents them in random order to make the selection strategy less apparent.

After obtaining the first n free choice decisions, SADR aims to identify the next n free choice questions that are most informative. Among the products presented, SADR calculates the difference between the predicted purchase probabilities of every two "consecutive" products (i.e., products ordered closest to each other) using a binary logit function. The most informative free choice questions are those for which the calculated difference is largest. For example, as in Figure 4, if the purchase probabilities for the 64th, 128th, and 192nd products are 1%, 10%, and 70%, respectively, the two consecutive products with the largest predicted difference between them are the 128th and 192nd products ($60\% = |70\% - 10\%|$). Dividing this range into four equal-sized ranges then leads SADR to choose products at positions 144, 160, and 176 for the subsequent iteration. The procedure terminates after it completes m iterations.

Section 1.3 of the online appendix provides recommendations about the parameterization of the SADR method, pertaining to its parameters n and m. The most recommended combination is $n \times m \geq 6$, $n > 1$, $m > 1$, and $n = m$. If $n = m$ is not possible, then $m > n$ is recommended, because it controls better for redundant free choice questions.

3.3. Scale-Extended Hierarchical Bayes Sampler to Estimate Final Utility Functions

Both SDR and SADR use a multinomial logit model in the hierarchical Bayesian sampler (for SADR, in step 6). They treat respondents' answers to the blocks of forced and free choice questions as two independent data sources, because Swait and Andrews (2003) argue that if two data sources rely on different elicitation procedures but capture the same utilities, their parameter values are equal, but only up to a certain scale. The forced choice decisions should provide information about the parameter value of each attribute level. The free choice decisions also provide some information about these parameter values, but their contribution mainly pertains to the utility relative to the no-purchase option (typically captured by the intercept). Both methods adopt the extended version of the hierarchical Bayes sampler used by Brazell et al. (2006), which explicitly allows the scale to differ between the forced and free choice decisions. The scale serves as a multiplier of the deterministic utility and is inversely related to error variance, which offers a measure of respondents' decision-making consistency across both data sources. Section 1.4 of the online appendix presents the details about the sampler.

4. Analysis of Endogeneity Concerns

Adaptive mechanisms that use a respondent's answers in previous questions to select additional questions might lead to endogeneity in the parameter estimates. Therefore, we investigate whether SADR suffers from such concerns.

From a statistical perspective, the adaptive mechanism might cause endogeneity because the commonly used log-likelihood functions for the estimation assume independence across choices. In SADR, questions at later stages in the free choice block depend on previous choices and thus on realizations of the error term. Yet, Liu et al. (2007) argue that the mechanism of creating $X_{\text{free choice}}$ can be ignored if the estimation adheres to the likelihood principle and if $X_{\text{free choice}}$ does not contain information beyond $Y_{\text{forced choice}}$ and $Y_{\text{free choice}}$. In a stylized example, they compare estimations of an adaptive method with the results of a nonadaptive (fixed design) variant and demonstrate that the results are not affected by the design-generation mechanism. We transfer this test of statistical bias to our proposed method and thereby demonstrate that SADR's estimated parameters also are unaffected (§6.1, online appendix).

In addition, SADR might suffer from behavioral endogeneity concerns, which might affect the estimation if the adaptive nature of the free choice block influences a respondent's choices while the estimation ignores this impact. The free choice block includes elements similar to another method, used extensively in economics: the double-bounded dichotomous choice method (DBDC; DeShazo and Fermo 2002). This method asks respondents two consecutive free choice questions: whether they would buy a certain product at a given price and then, in an adaptive and systematically varying method, at a higher or lower price, depending on whether the respondents stated that they would buy or not.

Researchers have studied whether the adaptive nature of these questions leads to anomalies in respondents' behavior. They document multiple possible explanations for the anomalies, such as anchoring, framing, or yea-saying (DeShazo and Fermo 2002). Alberini et al. (1997) find structural downward shifts, in that the follow-up (free choice) question lowers sample-wide average WTP. To overcome this bias toward decreasing WTP, Veronesi et al. (2011) propose using a well-balanced, symmetric design, which systematically varies the starting prices, because such designs should yield more modest biases, even if anchoring is strong.

However, instead of varying just the price, as in DBDC, SADR varies all attributes of a product in the subsequent questions. Therefore, anchoring effects are less likely. In addition, by making the number of free choice questions in each iteration greater than 1 ($n > 1$) and randomizing their order, we obtain a well-balanced, symmetric design, similar to that proposed by Veronesi et al. (2011). Randomizing also ensures that the selection strategy is less apparent. We adopted various tests from the DBDC literature to identify any structural shifts in WTP over later choice sets in the SADR results; no such shifts were observed in any of the three empirical studies (§6.2, online appendix). Thus, SADR appears robust to statistical and behavioral endogeneity concerns.

5. Monte Carlo Simulations to Compare Separated Adaptive Dual Response (SADR) and Separated Dual Response (SDR) with Choice-Based Conjoint and Dual Response

We start by summarizing the results of a first Monte Carlo simulation study, which we describe in more detail in the online appendix and for which we sampled 1,680 sets of individual parameters, representing the preferences of 100 respondents each. This simulation compares SADR and SDR with choice-based conjoint and dual response in terms of how well they recover the parameters of the final utility function and WTP. With half as many free choice questions compared with dual response, SADR recovers the parameters just as well as dual response and substantially better than choice-based conjoint. In addition, SADR does not suffer accuracy decrements, compared with other methods, when it includes more attributes and therefore has fewer degrees of freedom. These simulations do not mimic behavioral effects by varying the positions of the questions, so the SDR results are identical to dual response.

Next, we consider whether the WTP estimates are better for respondents with a lower, relative to a higher, purchase probability. The share of no-purchase decisions reflects purchase probabilities, such that a high share (here, approximately 50% in the choice-based conjoint condition) indicates a low probability, and a low share (here, approximately 10%) indicates a high probability. Table 1 compares the root mean squared error (RMSE) of the constant of the final utility function across different methods (§2, online appendix). We used the constant for the comparison because it serves as an anchor point; in cases in which a respondent chooses a given product over the no-purchase option, an accurate constant shifts the deterministic utility for that product above the deterministic utility associated with the no-purchase option (often treated as 0) and below the deterministic utility of the no-purchase option otherwise.

In choice-based conjoint and dual response, the constant for respondents with a low probability of buying gets captured significantly better than does the constant for respondents with a high probability ($p < 0.01$). Specifically, for dual response, the RMSE difference is 53% ($(0.89 - 0.58)/0.58$). The various SADR configurations instead yield similar RMSE values, regardless of respondents' purchase probabilities. The best performance comes from SADR[$m = 4$, $n = 4$], with two fewer free choice questions than required for choice-based conjoint or dual response. Thus, we can conclude that the use of the two new features in SADR leads to very good recovery of both the parameters and WTP.

Table 1. Monte Carlo Simulation Study: Ability to Recover Constant of Final Utility Function (RMSE)

Purchase probability	Share of no-purchases	Choice-based conjoint	Dual response and SDR	SADR [$m = 1$, $n = 9$]	SADR [$m = 9$, $n = 1$]	SADR [$m = 3$, $n = 3$]	SADR [$m = 2$, $n = 2$]	SADR [$m = 4$, $n = 4$]
Low	High	0.56	0.58	0.71	0.70	0.70	0.83	0.61
Middle	Middle	0.63	0.68	0.73	0.71	0.72	0.85	0.62
High	Low	0.82	0.89	0.74	0.72	0.73	0.86	0.63
Mean		0.67	0.72	0.73	0.71	0.72	0.85	0.62

Notes. RMSE, root mean squared error; lower values indicate better ability. SDR, separated dual response; SADR, separated adaptive dual response; m = number of iterations in free choice block, each of which consists of n free choice questions.

6. Three Empirical Studies

6.1. Tablet Study: Comparison of Separated Adaptive Dual Response (SADR) and Separated Dual Response (SDR) with Existing Discrete Choice Experiments

In our first empirical study, we examine the performance of the two new features implemented in SDR and SADR. Overall, we compare seven types of discrete choice experiments that gradually consider all characteristics described in Figure 1: all five methods listed in Figure 1 but three differently parameterized versions of SADR. The first version of SADR uses [$m = 1$, $n = 12$] and therefore the same number of free choice questions as in SDR, but it adaptively selects products for these free choice questions based on the observations collected in the forced choice questions. The second version of SADR uses [$m = 1$, $n = 6$], and the third one uses [$m = 3$, $n = 2$]. The latter two versions require only half as many free choice questions as SDR. In addition, SADR[$m = 3$, $n = 2$] extends the adaptive selection strategy by using decisions in previous free choice questions to select products for the next n free choice questions. We start by examining whether choice behavior with respect to the free choice questions differs across methods. Then, we compare internal and predictive validity, the resulting WTP estimates, and the impact on price demand functions.

6.1.1. Study Design. The study topic was tablets, with the following attributes and levels: brand (Apple, Samsung, Smarttab), display size (7", 10"), battery capacity (7 h, 11 h), display resolution (1,280 × 800px, 2,560 × 1,600px), storage capacity (16 GB, 32 GB, 64 GB), and price (€100, €250, €400, €550). We informed respondents that each tablet could access the Internet through wireless local area network (WLAN) and omitted 3G and LTE card slots as potential attribute levels, because they currently play a minor role in the tablet market. We also omitted weight as an attribute, to avoid correlations with display size.

To ensure realistic market prices, we added a brand-specific base price to the systematically varied product prices. These base prices were €0 for the fictitious brand Smarttab, €50 for Samsung, and €150 for Apple. Adding these base prices matches the recommendation to select prices in such a way that their ranges are slightly greater than in reality, here by approximately ±€20–30 (Rao 2014, p. 44).

We randomly assigned each respondent to one of the seven experimental treatments and used the same 12 choice sets in all treatments. Each choice set consisted of 3 product alternatives (plus the no-purchase option in choice-based conjoint and IACBC). Thus, we collected 36 observations to estimate nine parameters for the linear probability model (i.e., constant, price parameter, two parameters each for brand and storage capacity, and one parameter for each of the remaining attributes). To avoid lexicographic decision making, we used a design that allowed for some overlap in the attribute levels. After the discrete choice experiment, we implemented the same predictive validity tasks across all discrete choice experiments, asking respondents to make decisions in two forced and two free choice questions. We implemented and executed all methods in study 1 and in later studies using the online survey platform DISE (Schlereth and Skiera 2012a).

6.1.2. Data Collection. The tablet study produced 1,450 completed questionnaires. We recruited respondents through different channels, such as postings in specialized forums and invitations to respondents to previous studies, who agreed to be notified about new studies. As an incentive, we offered entry in a lottery for a tablet of the winner's choice and 10 gift vouchers valued at €20 each.

For IACBC, we excluded 25 of the 223 respondents because they were exposed to tablets with prices higher than €10,000. As these high values indicate, an analyst using IACBC can lose control of the range of reasonable prices. In total, 1,425 respondents entered the comparison of the different experimental treatments.

Table 2 summarizes the results, which we discuss in detail next and in §3 of the online appendix. According to the mean t-test comparisons, the subsamples did not differ significantly in their demographic and socioeconomic information ($p > 0.10$).

6.1.3. Purchase Decision Behavior. One of the aims of SADR is to avoid extreme responses (i.e., always or never choosing the no-purchase option); 94.2% of the respondents in SADR[$m = 3$, $n = 2$] stated at least once that they would purchase a product and also chose the no-purchase option at least once. In contrast, 36.0% of respondents in the choice-based conjoint and 22.2% of respondents in the dual response conditions exhibited extreme response behavior, such that their responses provided no information about when they would start or stop buying the product. Two discrete choice experiments performed very well in avoiding extreme response behavior: IACBC with only 1.0% of extreme responses and SDR with only 7.3%. The reason for IACBC's low share of extreme responses is that the prices varied widely, from €0.07 to €9,973, and thus exceeded the range in the other questionnaires (between €100 for Smarttab and up to €700 for Apple). The low share of SDR in terms of extreme responses, relative to dual response, is remarkable, because both methods ask the same free choice questions. We thus find evidence that the dual response sequences (forced choice question followed by a free choice question) induce choice deferral behavior. In support of this finding, the total share of no-purchase decisions in dual response is 46.2%, more than double the share observed in choice-based conjoint (22.7%). For SDR, the

Table 2. Tablet Study: Summary of Empirical Results

	Choice-based conjoint	Improved adaptive choice-based conjoint[a]	Dual response	Separated dual response	Separated adaptive dual response		
					$[m=1, n=12]$	$[m=1, n=6]$	$[m=3, n=2]$
Purchase decision behavior	$N=214$	$N=198$	$N=203$	$N=219$	$N=189$	$N=194$	$N=208$
Share of decisions not to purchase (%)	22.7	46.2	46.2	42.2	—	—	—
Share of respondents always purchasing (%)	32.7	0.0	9.4	3.7	0.0	0.5	0.5
Share of respondents never purchasing (%)	3.3	1.0	12.8	3.7	5.3	16.5	5.3
Share of extreme response behavior (%)	64.0	99.0	77.8	92.7	94.7	83.0	94.2
Model fit and parameter values							
Log-marginal density—Base model	−1,317	−981	−1,737	−2,090	−1,572	−1,303	−1,455
Log-marginal density—Scale model			−1,728	−2,073	−1,554	−1,325	−1,432
Relative scale μ_2/μ_1	—	—	1.03 (0.11)	1.24 (0.11)	1.28 (0.12)	0.95 (0.13)	1.36 (0.22)
Internal validity (12 choice sets)							
First choice hit rate (%)	90.5	91.5	90.6	88.9	89.2	91.9	91.1
Mean absolute deviation	0.19	0.16	0.17	0.19	0.18	0.15	0.16
Correlation of preliminary preference order with order based on final estimates	—	—	—	0.93	0.91	0.94	0.93
Predictive validity (2 forced choices)							
First choice hit rate (%)	73.6	64.4	75.9	76.3	71.4	76.0	77.2
Mean absolute deviation	0.32	0.38	0.30	0.30	0.34	0.30	0.28
Predictive validity (2 free choices)							
First choice hit rate (%)	73.1	79.0	77.6	81.7	81.2	80.4	81.7
Mean absolute deviation	0.28	0.23	0.26	0.23	0.23	0.24	0.22
WTP estimation comparison[b]							
Mean WTP[b]	€620	€549	€408	€362	€445	€444	€369
Median WTP[b]	€514	€497	€316	€373	€400	€434	€348
Survey characteristics							
Number of separate free choice questions (considered alternatives)	— (48)	— (48)	12 (60)	12 (60)	12 (60)	6 (48)	6 (48)
Share of redundant free choice questions (%)	—	—	67.7	65.0	63.0	45.9	35.2

[a]We removed 8.1% of respondents who saw prices greater than an arbitrary threshold of 10,000€. Prices originally ranged from €0.05 to €819,250.
[b]WTP estimation for a Samsung, 32 GB, 11 h, 10″, 2,560 × 1,600px display.

total share of no-purchase decisions is slightly lower than but still close to that in dual response (42.2% versus 46.2%). We did not compare this metric for SADR, because respondents assessed different attribute-level combinations in the free choice questions.

By comparing the differently parameterized versions of SADR, we learn more about how parameterization affects extreme response behavior. All three versions of SADR included almost no respondents who indicated that they would always make a purchase. Yet, in SADR$[m=1, n=6]$, the share of respondents who always selected the no-purchase option was approximately three times higher than that in SADR$[m=1, n=12]$. Fewer free choice questions can increase the likelihood of extreme response behavior. For example, if we consider the first six free choice questions in SADR$[m=1, n=12]$, we observe that 11.6% of respondents always defer purchases. We used logistic regression to test whether the position of the product alternative in the preference order could explain respondents' greater likelihood of always selecting the no-purchase option, but we found no significant effect ($p \gg 0.1$). Introducing the adaptive mechanism with respect to previous free choice decisions in SADR $[m=3, n=2]$ lowered the share of respondents who showed extreme response behavior to a level comparable with that of SADR$[m=1, n=12]$.

6.1.4. Estimation and Test for Endogeneity Concerns. To estimate preferences, we used a hierarchical Bayesian sampler (see §1.4, online appendix). If supported by improvements to the log marginal density, we used the scale-extended model, which treats the forced and free choice questions as two separate data sources and accounts for differences in error variances.

We also examined, for all adaptive methods, whether behavioral concerns of endogeneity are justified. In all three versions of SADR, we do not observe a significant shift in total utility or WTP in later free choice questions; that is, we find no indications of behavioral concerns of endogeneity. In contrast, for IACBC, we observe a significant negative shift in utility that indicates that IACBC's measured WTP decreases with later choice sets, indicating justifiable behavioral concerns of endogeneity. We provide the detailed results in the online appendix, §§6.2 and 6.3.

6.1.5. Internal and Predictive Validity. Because SADR uses different products for each respondent in the free choice questions, we compared the various methods in terms of the internal validity of the forced choice decisions, using the first choice hit rates and mean absolute deviation. Internal validity was high for all methods. Respondents were consistent in their decision making, likely because of their well-defined, consistent preferences. In particular, brand polarized these respondents into groups of Apple versus Android fans. For example, 10.9% of respondents always picked Apple, if possible.

To assess the goodness of the linear probability model as another measure of internal validity for SDR and SADR, we calculated the correlation between the preliminary preference order derived from the linear probability model and the preference order derived from the final parameter estimates. The Spearman rank correlation coefficient was at least 0.91, which is very high.

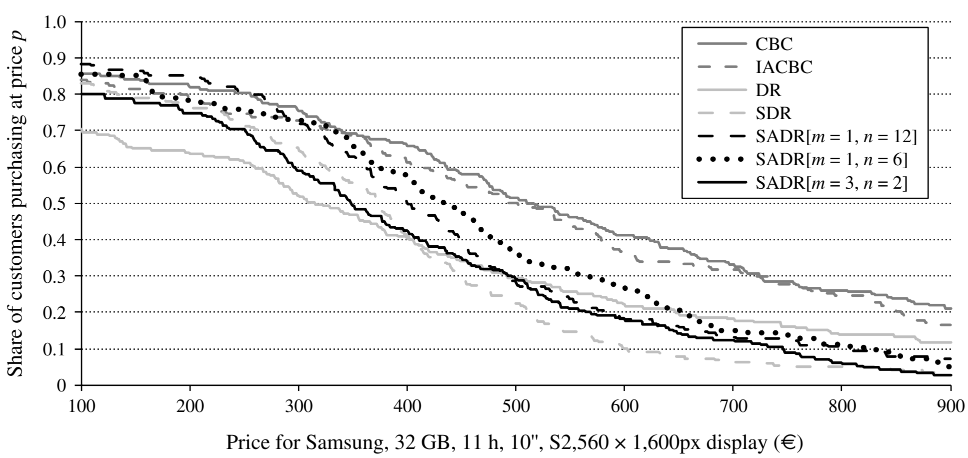
The predictive validity with respect to the forced choice questions was best for SADR[$m = 3$, $n = 2$] and worst for IACBC. Similarly, IACBC performed slightly worse than choice-based conjoint in Gensler et al. (2012). This result is not surprising, in that IACBC's adaptive price-setting approach creates situations in which respondents pick the no-purchase option in approximately half of the choice sets, so only the other half of choice sets can serve to make predictions about respondents' preferences.

The predictive validity with respect to the free choice questions was best for SDR and SADR[$m = 3$, $n = 2$]. Again, the differences across methods were small, because of the consistency in respondents' choices, but we observe that SDR and all versions of SADR easily outperformed the traditional methods such as choice-based conjoint or dual response in predicting when respondents would stop or start making a purchase.

6.1.6. WTP and Price Demand Functions. To augment the internal and predictive validity analyses, we compared the mean and median WTP estimates for an exemplary product derived from the estimates of each method. Without loss of generality, the chosen product had the following features: Samsung, 32 GB, 11 h, 10", 2,560 × 1,600px display. Choice-based conjoint provided the highest mean and median WTP, a consequence of the 32.7% of all respondents who would purchase in all choice sets. However, IACBC's mean and median WTP also were relatively high. The lowest median WTP estimates came from dual response, 38% lower than those in choice-based conjoint (i.e., |€316 − €514|/€514), because of the high share of no-purchase options selected. Both SDR and SADR[$m = 3$, $n = 2$] delivered about the same mean and median WTP. Finally, SADR[$m = 1$, $n = 12$] and SADR[$m = 1$, $n = 6$] resulted in nearly identical WTP estimates, higher than those of SADR[$m = 3$, $n = 2$].

By simulating demand functions (e.g., Miller et al. 2011, Wertenbroch and Skiera 2002), as in Figure 5, we provide visual comparisons of the substantive results of the seven discrete choice experiments, to clarify the differences in predicted market shares across the three methods.

Figure 5. Tablet Study: Comparison of Demand Functions



The visual comparisons in Figure 5 also help to distinguish managerial recommendations obtained from each method. For the high-price region, all SDR and SADR demand functions are low compared with the other discrete choice experiments, whereas demand is relatively high in the low-price region. Therefore, the managerial recommendations will be more conservative than those produced by choice-based conjoint, dual response, or IACBC. Choice-based conjoint and IACBC consistently overestimate demand for higher prices, compared with the other methods, so their managerial recommendations tend to encourage the highest prices; both methods predict a smaller decrease in demand. In this study, however, the high share of no-purchase decisions in the dual response conditions, such that 12.8% of respondents would never buy, resulted in the lowest market share predictions for low prices. For high prices, 9.4% of respondents would always buy in the dual response conditions, leading to higher market share predictions in high-price regions compared with those of SDR and SADR.

6.1.7. Redundant Purchase Decisions. In an ex post analysis, we also examine whether all free choice questions are necessary or if some of them might be redundant. A free choice question is redundant if the respondent already has answered about purchasing (not purchasing) a product with lower (higher) utility. Although we do not aim to eliminate all redundant forced choice questions (they provide information about the consistency and variance of unobserved behavior in free choice decisions), their number should be small. As we show in Table 2, the number of redundant free choices in SADR[$m = 3$, $n = 2$] was 32.5 percent points (= 67.7% – 35.2%) lower than the number in dual response.

Our experimental setup enables us to determine which features of SADR best reduce the share of redundant purchase decisions. That is, both dual response and SDR have the highest shares, but introducing the adaptive mechanism based on the forced choice questions, as in SADR[$m = 1$, $n = 12$], barely reduces it. Two changes were almost equally effective in reducing the share of redundant purchase decisions: First, asking only half of the free choice questions, as in SADR[$m = 1$, $n = 6$], reduced redundancy by 17.1 percent points (= 63.0% – 45.9%). Second, incorporating the adaptive component into the free choice block, by setting $m > 1$, reduced the percentage of redundant free choice questions by another 10.7 percent points to 35.2%. Therefore, the inclusion of multiple subblocks represents an important component of SADR.

In sum, the separation of forced and free choice questions into two blocks is the major driver of reduced extreme response behavior, resulting in substantially improved predictive validity. The introduction of the adaptive mechanism of selecting the free choice questions with $m = 1$ did not strongly improve predictive validity. However, the adaptive mechanism ($m > 1$) is highly effective and reduces the number of redundant free choice questions by 50%.

In IACBC, an analyst can lose control of the range of reasonable prices, and the predictive validity is rather low. Furthermore, we observe structural downward shifts in WTP for IACBC, justifying the behavioral concerns about endogeneity for IACBC.

6.2. Basketball Study: Comparison of SADR with Choice-Based Conjoint and Dual Response

The preceding results show that SDR and SADR perform better than choice-based conjoint or dual response in terms of predictive validity and reducing extreme response behavior. The results also demonstrate substantial differences in managerial implications, exemplified by the differences in the price demand functions. We examine the generalizability and robustness of these results in an additional empirical study, for which we cooperated with a major league basketball team to compare the predictions of the different discrete choice experiments. We recruited 459 fans to participate in a survey about their preferences for different kinds of home game tickets. Links to the surveys appeared in digital newsletters, on the basketball team's Facebook page, and websites. Approximately 92% of the respondents had attended at least one game in the previous season. Team management incentivized participation by raffling basketball shirts and free tickets. We compare SADR, as the most elaborated method, with choice-based conjoint and dual response.

6.2.1. Study Design. We identified relevant attributes and levels through pretests and discussions with the executive directors of the cooperating basketball team. The attributes were the price category (1 [base price: €27], 2 [base price: €22], 3 [base price: €17], 4 [base price: €12]), additional features (nothing, free parking, free VIP parking, free public transportation), and deviation from the base price (€ – 3, €–1, €1, €3). The 12 choice sets each contained representatives of all four price categories, such that we collected 48 observations to estimate 8 parameters for the linear probability model (i.e., constant, price parameter, 3 parameters for price category, and 3 parameters for additional features). For SADR, we used the configuration [$m = 3$, $n = 2$].

Table 3 summarizes the results, and §4 of the online appendix provides additional insights. According to mean t-test comparisons, the subsamples did not differ significantly in their demographic or socioeconomic information ($p > 0.10$).

6.2.2. Purchase Decision Behavior. As in the previous study, SADR avoided extreme responses: 86.71% of respondents stated at least once that they would purchase a product and chose the no-purchase option at least once. In contrast, 56.25% of respondents in the choice-based conjoint and 25.68% of those in the dual response conditions exhibited extreme response behavior, providing no information about when they would start or stop buying the product. Compared with the use of choice-based conjoint, using dual response increased the total share of no-purchase decisions by 13.56 percentage points, which nearly doubles the share observed in choice-based conjoint.

6.2.3. Internal and Predictive Validity and a Test for Behavioral Concerns of Endogeneity. Consistent with the results of the tablet study, internal and predictive validity was best for SADR. For SADR, we did not find that the constant of the (final) utility function changed with later free choice questions, so concerns about behavioral endogeneity did not seem justified (§6.2, online appendix).

6.2.4. External Validity. In Table 3, we also analyze external validity across the methods by comparing our predictions of shares of customers in each price category with the observed shares for the seasons before and after the study. Each season included 17 home games, and fans could choose among four price categories (we excluded VIP tickets and seats for disabled persons). According to managers, the team's marketing efforts and performance (i.e., ranking in the league) were approximately the same across the two seasons. We also could ignore restrictions due to capacity constraints, because tickets in the four price categories rarely sold out. In a second comparison, we exploited an actual change, in that between the two seasons, management raised prices by €2.

The RMSE between the predicted and observed shares of the price category ticket purchases was low for all three methods, ranging between 0.059 and 0.102 for the first comparison and 0.034 and 0.041 for the second. Still, the SADR consistently achieved the lowest RMSE. Choice-based conjoint predicted the lowest drop in attendance due to the price increase, of 1.59%; SADR and dual response provided similar predictions of 2.63% and 2.82% on average, closer to the observed drop of 4.36%. Still, SADR offered a small advantage, because its RMSE in the four price categories was lowest.

Table 3. Basketball Study: Summary of Empirical Results

	Choice-based conjoint	Dual response	Separated adaptive dual response [$m = 3$, $n = 2$]
Purchase decision behavior	$N = 160$	$N = 146$	$N = 153$
Share of decisions not to purchase (%)	15.83	29.39	—
Share of respondents always purchasing (%)	56.25	25.00	9.49
Share of respondents never purchasing (%)	0.00	0.68	3.80
Share of respondents without extreme response behavior (%)	43.75	74.32	86.71
Model fit and parameter values			
Log-marginal density—Base model	−1,261	−1,734	−1,550
Log-marginal density—Scale model	—	−1,735	−1,546
Internal validity (12 choice sets)			
First choice hit rate (%)	67.9	73.6	74.8
Mean absolute deviation	0.38	0.38	0.36
Correlation of preliminary preference order with order based on final estimates	—	—	0.81
Predictive validity (2 choice sets)			
First choice hit rate (only forced choices) (%)	65.2	71.2	74.2
Mean absolute deviation (only forced choices)	0.41	0.37	0.36
External validity			
RMSE of choice proportions in four price categories; season before study	0.099	0.095	0.059
RMSE of choice proportions in four price categories; season after study	0.102	0.098	0.061
RMSE of change in number of viewers per price category after price increase	0.041	0.038	0.034
Survey characteristics			
Number of separate free choice questions (considered alternatives)	— (60)	12 (72)	6 (60)
Share of redundant free choice questions (%)	—	67.8	42.3

6.3. Video-on-Demand Study: Replication of the Comparison of SADR with Choice-Based Conjoint and Dual Response

We conducted a third empirical study in a video-on-demand context, in cooperation with a German provider. In this study, we tested preferences for various bucket pricing plans for a video-on-demand service and used a nonlinear (final) utility function for the estimation (Schlereth and Skiera 2012b). The findings reaffirmed the results of the basketball and tablet studies, and we report them in detail in §5 of the online appendix.

In sum, extreme response behavior differed strongly across the methods: SADR had the lowest share (8%), whereas for choice-based conjoint approximately 17% always purchased and another 17% never purchased; for dual response over 30% never purchased and only 7% always purchased. Consequently, choice-based conjoint predicts unrealistic high mean, medium, and maximum WTP and dual response's mean and medium WTP predictions are substantially lower than those of SADR. Accounting for differences in the consistency between the forced and free choice questions using our scale-extended hierarchical Bayes sampler provided substantial improvements for both SADR and dual response. SADR was superior with respect to internal and predictive validity; it also surpassed substantially choice-based conjoint and dual response in external validity, measured here as the RMSE and differences in choice shares between the predictions from the discrete choice experiments and the self-stated past usage behavior.

7. Conclusions

In this paper, we propose two discrete choice experiment features that are especially helpful for estimating WTP or predicting demand: strictly separating all forced choice questions from all free choice questions, and using an adaptive mechanism to select fewer, but more informative, free choice questions. The use of the separation feature results in the proposed separated dual response (SDR) method, and the use of both (separation and adaptive selection) features produces the separated adaptive dual response (SADR) approach.

The purpose of these two features is to overcome shortcomings typically associated with the inclusion of the no-purchase option in discrete choice experiments, such as context effects with respect to the no-purchase option, extreme response behavior, and a lower ability to capture the WTP of respondents who exhibit a high probability of buying (compared with respondents who have low probability). Previous literature already has detailed the first two problems; the recognition and effort to address the third problem constitutes a key contribution of this paper.

The separation feature ensures that context effects with respect to the no-purchase option are not a concern, because the forced choice questions do not contain a no-purchase option, and all free choice questions contain only one product alternative, disconnected from the forced choice questions. In addition, because of the strict separation of forced and free choice questions, both SDR and SADR avoid choice deferral, a major problem for dual response experiments. The combination of both the separation and adaptive selection features with having only one product alternative in each free choice question also increases our ability to capture the WTP of respondents who have a high probability of buying (see §5). Additionally, we circumvent extreme response behavior by using a new adaptive procedure to select fewer, but more informative product alternatives in the free choice questions.

In three empirical studies, we demonstrate that SADR outperforms choice-based conjoint in nearly all dimensions: it substantially improves internal, predictive, and external validity; its accuracy in parameter recovery is substantially better; and it does not require additional cognitive effort. Nor does the adaptive nature of SADR create endogeneity problems. In our tablet study (§6.1), we show that the strict separation of forced and free choice questions into two blocks is the main driver of the reduction in extreme response behavior, resulting in substantially improved predictive validity. Because SDR and SADR performed nearly equally well, we recommend further studies that compare the methods empirically and more intensively. Introducing the adaptive mechanism of selecting the free choice questions did not strongly improve predictive validity, but the adaptive mechanism reduces the number of redundant free choice questions by nearly 50%. In addition, SADR obtains the best predictive validity results, with only half of the free choice questions.

Finally, SADR provides a new perspective on the market, and it yields managerial recommendations for price–demand relationships that differ from those derived from choice-based conjoint or dual response. The managers of the companies that cooperated with these studies considered the recommendations of SADR more valid. Thus, we conclude that the two features in SDR and SADR create attractive alternatives to traditional methods for measuring preferences and WTP.

Acknowledgments

The authors gratefully thank Maike Mohr, Jan-Michael Becker, Franziska Völckner, and Henrik Sattler for support of the execution of the tablet study. For help on the basketball study, the authors thank Christian Hinz and Marcus Nixdorf from Fraport Skyliners. The authors also thank Christian Strominski for his support of the video-on-demand study.

The authors give many thanks to Thomas Otter for his recommendations and especially for his help on the test of statistical endogeneity. The authors also give many thanks to the participants of the research seminar series and congresses at Cornell University (in particular the organizer, Vithala Rao), Monash Business School (Peter Danaher, Harmen Oppewal), University of St. Gallen (Andreas Herrmann, Reto Hofstetter), and University of New South Wales (Ashish Sinha), as well as University of Vienna (Udo Wagner). The authors also gratefully thank department editor Pradeep Chintagunta, the two anonymous referees, and the associate editor for their encouraging and constructive feedback. Last but not least, the authors thank Elisabeth Nevins Caswell for copyediting the paper.

Endnotes

[1] In our paper, WTP refers to the price for a whole product at which a respondent is indifferent between buying and not buying. This definition differs from those that refer to the price, given an attribute improvement (not the whole product), that leaves a consumer indifferent. We consider that value the equalization price instead.

[2] A demonstration survey, http://www.dise-online.net/demo.aspx, of the online survey platform DISE provides examples of all methods listed in Figure 1, in particular SDR and SADR.

[3] A Monte Carlo simulation study in §2 of the online appendix reveals how well the resulting preliminary preference order approximates the true preference order: the average Pearson correlation coefficient was 0.78.

References

Alberini A, Kanninen B, Carson RT (1997) Modeling response incentive effects in dichotomous choice contingent valuation data. *Land Econom.* 73(3):309–324.

Bech M, Kjaer T, Lauridsen J (2011) Does the number of choice sets matter? Results from a Web survey applying a discrete choice experiment. *Health Econom.* 20(3):273–286.

Brazell JD, Diener CG, Karniouchina E, Moore WL, Séverin V, Uldry P-F (2006) The no-choice option and dual response choice designs. *Marketing Lett.* 17(4):255–268.

DeShazo J, Fermo G (2002) Designing choice sets for stated preference methods: The effects of complexity on choice consistency. *J. Environ. Econom. Management* 44(1):123–143.

Dhar R (1997) Consumer preference for a no-choice option. *J. Consumer Res.* 24(2):215–231.

Dhar R, Simonson I (2003) The effect of forced choice on choice. *J. Marketing Res.* 40(2):146–160.

Feit EM, Beltramo MA, Feinberg FM (2010) Reality check: Combining choice experiments with market data to estimate the importance of product attributes. *Management Sci.* 56(5):785–800.

Gensler S, Hinz O, Skiera B, Theysohn S (2012) Willingness-to-pay estimation with choice-based conjoint analysis: Addressing extreme response behavior with individually adapted designs. *Eur. J. Oper. Res.* 219(2):368–378.

Heckman JJ, Snyder JM (1997) Linear probability models of the demand for attributes with an empirical application to estimating the preferences of legislators. *RAND J. Econom.* 28(0): 142–189.

Liu Q, Otter T, Allenby GM (2007) Investigating endogeneity bias in marketing. *Marketing Sci.* 26(5):642–650.

Louviere J, Hensher D, Swait J (2000) *Stated Choice Methods: Analysis and Application* (Cambridge University Press, Cambridge, UK).

Miller KM, Hofstetter R, Krohmer H, Zhang ZJ (2011) How should we measure consumers' willingness to pay? An empirical comparison of state-of-the-art approaches. *J. Marketing Res.* 48(1):172–184.

Parker J, Schrift R (2011) Rejectable choice sets: How seemingly irrelevant no-choice options affect consumer decision processes. *J. Marketing Res.* 48(5):840–854.

Rao V (2014) *Applied Conjoint Analysis* (Springer, Heidelberg, Germany).

Rooderkerk RP, Van Heerde HJ, Bijmolt THA (2011) Incorporating context effects into a choice model. *J. Marketing Res.* 48(4): 767–780.

Schlereth C, Skiera B (2012a) DISE: Dynamic intelligent survey engine. *Quantitative Marketing and Marketing Management* (Springer Gabler, Wiesbaden), 225–243.

Schlereth C, Skiera B (2012b) Measurement of consumer preferences for bucket pricing plans with different service attributes. *Internat. J. Res. Marketing* 29(2):167–180.

Swait J, Andrews RL (2003) Enriching scanner panel models with choice experiments. *Marketing Sci.* 22(4):442–460.

Toubia O, Hauser J, Garcia R (2007) Probabilistic polyhedral methods for adaptive choice-based conjoint analysis: Theory and application. *Marketing Sci.* 26(5):596–610.

Veronesi M, Alberini A, Cooper JC (2011) Implications of bid design and willingness-to-pay distribution for starting point bias in double- bounded dichotomous choice contingent valuation surveys. *Environment. Resource Econom.* 49(2):199–215.

Wertenbroch K, Skiera B (2002) Measuring consumer willingness to pay at the point of purchase. *J. Marketing Res.* 39(2):228–241.

Wlömert N, Eggers F (2016) Predicting new service adoption with conjoint analysis: External validity of BDM-based incentive-aligned and dual-response choice designs. *Marketing Lett.* 27(1):195–210.

Yu J, Goos P, Vandebroek M (2011) Individually adapted sequential Bayesian conjoint-choice designs in the presence of consumer heterogeneity. *Internat. J. Res. Marketing* 28(4):378–388.